DOCUMENT RESUME

ED 414 463 CE 075 383

AUTHOR	Taylor, Maurice
TITLE	Test Review: Workplace Assessment Tools = Examen de divers tests: Outils d'evaluation en milieu de travail.
INSTITUTION	National Literacy Secretariat, Ottawa (Ontario).
PUB DATE	1997-04-00
NOTE	102p.
AVAILABLE FROM	National Literacy Secretariat, Human Resources Development Canada, Ottawa, Ontario K1A 1K5, Canada; phone: 819-953-5280.
PUB TYPE	Guides - Non-Classroom (055) -- Multilingual/Bilingual Materials (171)
LANGUAGE	English, French
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Adult Basic Education; Adult Literacy; Basic Skills; Competency Based Education; *Criterion Referenced Tests; Foreign Countries; French; Job Skills; Job Training; Lifelong Learning; Literacy Education; *Minimum Competency Testing; *Norm Referenced Tests; Scores; Scoring; *Standardized Tests; Test Construction; Test Content; Test Reviews; *Test Selection; *Workplace Literacy

ABSTRACT

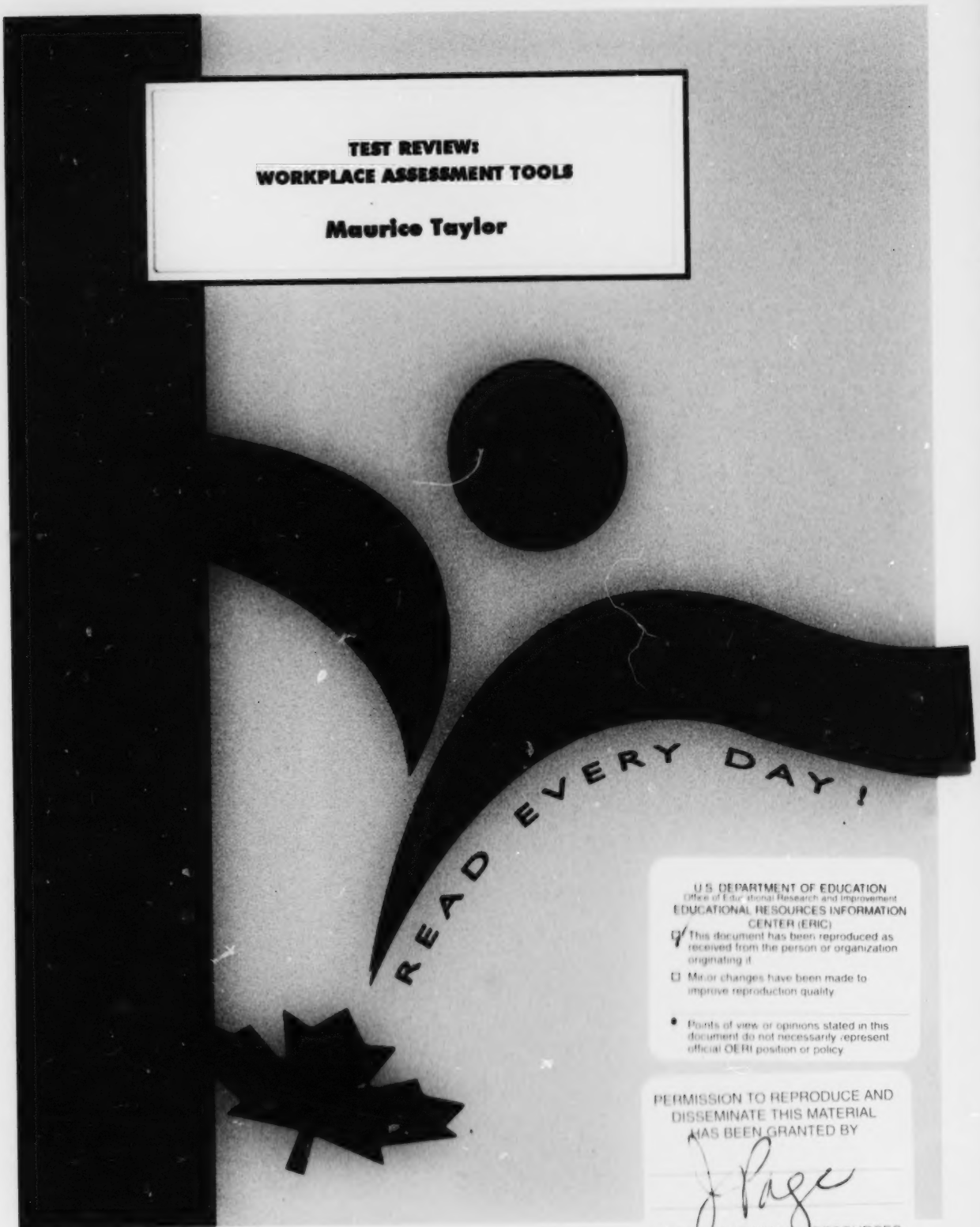
This publication (in both English and French versions) on learner assessment in workplace literacy programs begins with a discourse on some dimensions of the testing question. It discusses three major approaches to assessment in the broader area of adult literacy that can shed light in the domain of workplace literacy: traditional, competency based, and profile. The categories of essential skills are described, as a general understanding of these different categories enhances decisions around what types of instruments to use. The categories are: traditional basic skills, thinking skills, teamwork, skills required for continuous learning, and computer skills. Next, the publication describes standardized tests (norm referenced and criterion referenced) and competency-based tests. The second part of the document discusses eight assessment tools that have been used in workplace programs: Tests of Adult Basic Education (TABE) Work-Related Foundation Skills; Tests of Applied Literacy Skills; Basic English Skills Test; Canadian Adult Achievement Test; Measurement, Reading, and Arithmetic; Wide Range Achievement Test 3; and Metric Skills Assessment. Each instrument is described in terms of the test content, the description of scores, and the test development process, along with the source's address and telephone number. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 463

TEST REVIEW:
WORKPLACE ASSESSMENT TOOLS

Maurice Taylor



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Page

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

National Literacy Secretariat

Le Secrétariat national à l'alphabétisation

Human Resources Development Canada / Développement des ressources humaines Canada

Canada

Canada

BEST COPY AVAILABLE

CE 075 383

TEST REVIEW:
Workplace Assessment Tools

Maurice Taylor

Partnerships in Learning
April 1997

Human Resources Development Canada
Développement des ressources humaines Canada

National Literacy Secretariat
Le Secrétariat national à l'alphabétisation

To order copies of this report, contact:

National Literacy Secretariat
Human Resources Development Canada
Ottawa, Ontario K1A 1K5
Telephone: (819) 953-5280
Fax: (819) 953-8076
TTD: (819) 953-2338
E-mail: nls@fox.nstn.ca

Table of Contents

Part 1 Dimensions of the Testing Question

Assessment in Learning 2

What are the Essential Skills of Today? 4

The Nature of Testing 8

Part 2 Technical Test Review

Tests of Adult Basic Education (TABE) Work-Related Foundation Skills 11

Tests of Applied Literacy Skills 16

Basic English Skills Test 22

Canadian Adult Achievement Test 27

Measurement, Reading and Arithmetic 32

Wide Range Achievement Test 3 36

Metric Skills Assessment 41

INTRODUCTION

Learner assessment in workplace literacy programs is still a much debated issue in the field of adult education. It continues to mean different things to different people. On the one hand, the debate can focus on the essential components of a standardized test, and on the other hand it can center on ways to obtain relevant information in order to make decisions. In this discussion paper, one part of the assessment question is addressed—that of testing. By means of a technical test review, seven assessment tools that have been used in workplace programs are discussed. Each instrument is described in terms of the test content, the description of scores and the test development process.

As a way of anchoring this review, the topic is introduced with a discourse on some of the dimensions of the testing question. These sections will serve as the backdrop for helping practitioners make decisions around testing when it is a necessary component of a workplace education program. The thrust of this paper is not to suggest that some tests are better than others but rather to argue that workplace educators need to find the test that is most appropriate to their learner and program requirements.

PART 1

DIMENSIONS OF THE TESTING QUESTION

Assessment in Learning

In workplace education programs, assessment can be viewed like a map. It helps people arrive at a destination. The map or learning plan guides the trainee to a destination which are the goals and outcomes of the program. People often use maps to find a desirable route to their destination—some want the fastest route while others enjoy a more scenic route with points of interest along the way. The same holds true for assessment. It should help learners move toward their goals and tell them how much time will be needed to reach their goals. In many of the current workplace education programs assessment is being used to place learners into levels, to prescribe specific learning plans, to monitor progress and to certify attainment and goal completion. Many practitioners and scholars have advocated that good assessment, curriculum and instruction are all part of a learning plan. This plan is made up of various strategies and techniques for instructing, assessing and reinforcing learning. Just as no single teaching strategy is effective for all learners in all situations, no single assessment

instrument or process is adequate in providing quality information for all purposes.

Generally, there are three major approaches to assessment in the broader area of adult literacy and these can shed some light in the domain of workplace literacy. Each approach carries certain assumptions and limitations. In the traditional approach, functional literacy is equated with the attainment of a particular grade level score on tests of achievement. Through the use of such tests, it is possible to estimate percentages of various population groups performing at or above specified reading grade levels. Persons performing at or above a specified level are considered to have adequate skills to perform successfully on materials or tasks judged to be of comparable grade level difficulty.

The second approach is called competency based. In this approach tasks are developed from non-school types of materials and administered to national samples. By analyzing task responses, it is possible to estimate the proportion of the adult population that can be expected to perform the tasks successfully and to determine the extent to which various background characteristics such as educational level, race, ethnicity, gender and income level are related to the estimates of successful performance.

In the profile approach the conception of literacy is not simply viewed as a set of isolated skills associated with reading and writing, but more importantly it is seen as the application of those skills for specific purposes and specific contexts. Because of the diversity of literacy demands that people encounter at home, at work and in their communities, literacy is seen as a set of context utilized practices rather than as a single standard. Three distinct aspects of literacy are included in this approach—prose, document and quantitative.

As the reader moves into the technical test review section, each one of these three approaches to assessment is represented.

What Are the Essential Skills of Today?

Another side of the assessment and testing issue is related to the types of skills required in the workplace. One factor in choosing an instrument is to consider whether or not the skills assessed on the test are the same kinds of skills being used in the job or in future jobs. This line of inquiry has been going on for some time now. Most recently the goal of the Essential Skills Research Project being conducted by Human Resources Development Canada was to find out

what workers do in their jobs. Through approximately 3,000 interviews with workers in a broad range of industries across Canada, the study examined the essential skills actually required in lower skill entry level occupations. The objective of the study was to produce for each of these occupations, a profile describing the essential skills required in the Canadian workplace. These profiles contain descriptions of how the essential skills are used in an occupation as reported by the actual job incumbents and include the more traditional basic skills, thinking skills, teamwork and skills required for continuous learning. Having a general understanding of these different categories of skills will enhance decisions around what types of instruments to use. The reader should note here that the seven tests later reviewed do not purport to measure all of the skill categories described in the following paragraphs.

For the traditional basic skills such things as the reading of textual materials, the use of documents, writing, numeracy and oral communication are included. The following are examples of some of those skill categories. For textual reading some of the typical and more complex tasks are rated on a scale ranging from reading relatively short texts to locating a single piece of information or following simple written instructions to the more complex task of interpreting dense and complex texts.

In the skill category—use of documents—there are typical and complex tasks which involve reading. These are rated on a scale of difficulty from reading signs, labels and other table like texts which require only identification of key words to making high level text-based inferences and using specialized knowledge. For the skill of writing there are typical and more complex tasks also rated on a scale with level 1 skills including such things as preparing written materials in preset formats requiring very brief text to the more complex task of preparing written materials on a diverse range of topics to persuade or present an analysis, comparison, evaluation or critique.

For numeracy, tasks are categorized by application setting: money math, scheduling or budgeting math, measurement in calculation math or data analysis math. Numeracy is further described by specific math skills such as whole numbers, fractions, proportions, measurement conversions and geometry. The skill of oral communication is described by type and purpose such as listening, speaking, interacting with co-workers, interacting with those who supervise, interacting with customers, participating in group discussions, presenting information to large groups and leading formal discussions.

The second major category of skills is called thinking skills with one of the first sets of sub-skills related to problem solving. Problem solving skills are also

rated and include such factors as complexity of the problem, complexity of identifying the problem, complexity of identifying the solution steps and complexity of assessing the situation. Related to problem solving is decision making. Typical and most complex tasks are again rated on a point scale and include such factors as whether there is a set of procedures or decision tree to follow, the consequence of error and the extent to which judgement is required to make an appropriate decision. The next set of sub-skills under thinking skills are planning and organizing one's own job tasks. The planning and organizing of job tasks is rated on a scale and includes such factors as the extent of variety in work activities to the extent to which the worker's own work plan must be integrated with the work plans of others. The next category is use of memory. The content is described by type, for example remembering information for brief periods of time, remembering unique events and remembering job specific information for long-term use.

The third major category of skills is working with others. Content is described by the work context in which the job is performed, for example working alone, independently, with a partner or a helper or in a team, and by participation in leadership or supervisory activities such as orientating new employees, assigning tasks to other workers, selecting contractors or suppliers or

participating in formal group discussions about work processes. Computer skills is the next category. The use of computers is rated on a five-point scale ranging from "uses computers but no knowledge of software required" to "assesses needs and designs computer systems".

In summary, essential skills are those skills that are found in virtually all occupational areas, though their specific form or level of complexity varies. These same skills provide the foundation for learning other more specific skills and enable an individual to perform the tasks required in the job.

The Nature of Testing

A common concern for many practitioners trying to make sound decisions about assessment is that the technical terminology of testing is cumbersome. Although reported in other sources, this part of the discussion paper reviews some of the terms related to testing. For example standardization, norms, absolute standard and competency-based testing are all terms used in the technical review section of the report. A standardized test is a test that is administered under standard conditions to obtain a sample of learner behaviour that can be used to make inferences about the learner's ability. A standardized test differs

from an informal test in that the latter does not follow a fixed set of conditions. For example in a standardized achievement test, the same reading materials are read by different learners following the same procedures, answering the same types of questions and observing the same time limits.

There are two types of standardized tests: norm referenced and criterion referenced. Many standardized tests have been adopted to permit a learner score to be interpreted in relation to or in reference to the scores of other people who have taken the test. This type of test is called norm referenced. In such a case an individual's standardized test score is interpreted by comparing it to how well the referenced group normally performs on the test.

The second type of standardized test is criterion referenced. This testing format is closely related to the development of self-paced, individualized program instruction. In programs following this approach, a domain of knowledge or skills is carefully defined. An absolute standard or criterion of performance is set and everyone's score is established in relation to that standard. Learning objectives that can be assessed are specified and the units of instruction frequently called modules are developed to teach the various subsets of knowledge and skill identified by the learning objectives.

Another type of test is a competency-based test. Typically, developers first identify the important objectives or competencies to be learned. Next, test items are developed to determine whether learners already possess the competencies or if instruction is needed to develop certain competencies. Then, various commercially available materials with a variety of learning exercises are identified that teach each of the competencies so that teachers can select the materials their learners need to master. This approach is a form of teaching to the test, even though the exact contents of the assessment instrument may not appear in the curriculum to avoid directly teaching to the specific test items. In the competency-based programs, the competencies are identified first and the curriculum is designed to help the learner achieve these specific competencies.

PART 2

TESTS OF ADULT BASIC EDUCATION (TABE)
Work-Related Foundation Skills

CTB MacMillen-MacGraw Hill
20 Ranch Road
Monterey, California
93940-5703
Tel.: 1800-538-9547

Test Content

The Work-Related Foundation Skills (WRFS) is a series of norm referenced tests designed to measure achievement of basic skills in reading, mathematics and language commonly found in a majority of adult basic education curricula. Using appropriate content and language to represent the workplace context, WRFS assesses achievement skills and individual needs required to function in the workplace and in society.

WRFS has four forms: health, trade-technical, business-office and general. The test items, graphics and passages are presented in a context of these workplace environments. All four forms of the WRFS measure comparable skills. Because the four forms are comparable in content coverage, they can be used effectively in a pre-test / post-test situation.

Each of the four forms of WRFS contain the following subtests: reading, mathematics computation, applied mathematics and language. It take about 30 minutes to complete each subtest. The reading subtest focuses on the learner's understanding of the reading process as well as on the ability to construct meaning from the text. In the context of passages, graphics and reference sources, this subtest measures the examinee's ability to recognize words in work milieus, recall information, and recognize main ideas. The reading items focus on functional reading through the use of diagrams, charts, forms and schedules.

The mathematic's computation subtest measures the basic computational skills necessary for any successful mathematical program. It measures the examinee's understanding of the operations of addition, subtraction, multiplication and division as they relate to whole numbers, decimals and fractions. Also measured is the student's understanding of money, integers and percents. The applied mathematics section of the test stresses concept formation, reasoning and problem solving in context. It measures the examinee's ability to apply mathematical concepts related to renumeration, computation in context, number theory, measurement, data interpretation, geometry and prealgebra. Attention is given to the reasoning skills required for practical problem solving.

The language subtest measures the learner's skills in language usage,

sentence formation, paragraph development and writing conventions such as capitalization and punctuation. The subtest emphasizes skills essential to the achievement of full literacy and stresses the link between language and reading and writing. It also places importance in using language properly in the workplace. A variety of authentic work related context are presented as stimulii.

Description of Scores

The WRFS provides norm referenced scores in the form of scale scores, grade equivalence, and percentile ranks. Additionally, it provides criterion reference information for objectives embedded in the tests. The WRFS subtest and objective structures were determined by compiling information from various state adult education curricula and by studying the outcomes and competency frameworks of adult programs throughout the United States. The WRTS standardization sample was collected in 1994 and invoked a nation wide sample of approximately 1700 adult learners. Each study participant took the TABE WRTS and Form 7 and 8.

Test Development

The creation of the TABE WRFS followed a typical test development process which included the following seven steps: assessment design, writing, item review, tryout, item analysis, test selection and norms development.

Reliability. The Kuder-Richardson Formula 20 coefficient based on the number of correct scores of the WRFS range from .82 - .91. The higher the KR 20 coefficient, the greater the internal consistency of the subtest level. Internal consistency indicates that each subtest level as a whole measures a single trait, such as reading ability rather than a combination of two or more abilities.

Validity. Content related validation in achievement tests is evident by a correspondence between test content and instructional content. To ensure such correspondence, the test developers conducted a comprehensive curriculum review and met with educational experts to determine common educational goals and the knowledge and skills emphasized in today's curricula. This information was used to guide all phases of the design and development of this work related

assessment tool.

Construct validity. Construct validity which refers to the meaning of test scores and the kind of inferences they support, is a central concept underlining the work related assessment test validation process. Evidence for construct validity is comprehensive and integrates evidence from both content and criteria related validity. For example, to demonstrate comprehensiveness, the work related assessment tool must contain items that represent essential instructional objectives. Additionally, patterns of inter-correlations among the work related assessments and between them and other instruments in the TABE family of assessments, demonstrate convergent and divergent validity. That is, tests which are designed to measure similar skills correlate more highly than tests designed to measure distinctly different skills. For example, the WRFS language subtest correlates highly with the reading subtest, as expected, while its correlation with the WRFS mathematic subtests are lower.

Tests of Applied Literacy Skills

Educational Testing Service
Language Learning and Assessment
Rosedale Road
Mail stop 05-9
Princeton, New Jersey 08541
Telephone: 1-800-551-1230

Test Content

The Tests of Applied Literacy Skills (TALS) measures three distinct and important aspects of literacy—prose, document and quantitative. Each of these three subtests is available in two forms. Prose literacy tasks involve the knowledge and skills needed to understand and use information from texts that include editorials, news stories, poems and the like. The two forms of the Prose Literacy Test contain a total of 48 tasks. The materials in the test are mostly expository, which means that they describe or define. The texts used in the test are reprinted in their entirety and replicate the layout and typography of the original sources. Three aspects of prose literacy are represented in this subtest: locating, integrating and generating information. Tasks from each of these three areas and the three subtests extend over a range of difficulty. A twenty minute time limit is specified for each of the two sections of each test form.

Document literacy tasks involve the knowledge and skills required to locate and use information contained in job applications, payroll forms, transportation schedules, maps, tables, indexes and so forth. The two forms of the Document Literacy Test contain a total of 52 tasks. Many of the tasks are a necessary part of meeting the requirements of a job and managing a household. These skills involve strategies needed to locate information in various complex arrays such as graphs and charts and to transfer this information from one document to another. Three types of questions or directives are used in the document scale: locating, cycling and integrating.

Quantitative literacy tasks involve the knowledge and skills needed to apply arithmetic operations, either alone or sequentially to numbers that are embedded in print materials such as balancing a checkbook, figuring out a trip, completing an order form or determining the amount of interest from a loan investment. There are 46 tasks in both forms of the Quantitative Literacy Test. The respondent must perform addition, subtraction, multiplication and division using numbers that are buried in printed materials encountered in everyday practical situations. The difficulty of the tasks are a function of the particular arithmetic operations called for, the extent the number is embedded in the material and the

extent to which an inference must be made to identify the type of operation to perform.

Description of Scores

The first step in interpreting the test results is to calculate the raw scores. The raw scores are the total number of right answers from each section of a particular form and are obtained by scoring the questions on a right/wrong basis using the guides in the test manual. The second step is to convert raw scores to proficiency scores using the tables in the manuals for each form of the three literacy tests. To obtain an individual's proficiency score the test administrator selects the appropriate table of the form of the test and locates the number correct for each of the two sections of the test. The number in the table that is defined by the intersection of the appropriate section rows is the proficiency score. This score provides an estimate of an individual's demonstrated skill within each of the literacy domains.

With a proficiency score in hand, the test administrator uses the interpretive framework in the manual which characterizes selected tasks at various levels of difficulty along each of the literacy scales. The scales are

designed to range from 0-500, however the majority of tasks fall between 200 and 400 on each scale. To understand what each test measures, the tasks are interpreted in light of the interactions between the stimuli and the questions or directive that can account for placement of tasks along each of the scales. The interpretive framework for the Prose Literacy Test focuses on locating information in the text, integrating information from text and generating information from text. Similarly, the interpretive framework for the Document Literacy Test includes locating information in documents, cycling through information in documents, and integrating information in documents. Conditions and levels of difficulty for the Quantitative Literacy Test center on addition, subtraction, multiplication and division tasks and combination tasks.

At the student level, the TALS can serve a variety of needs. Each of the tests provide an estimate of an individual's demonstrated proficiency in applying a set of literacy skills to selected materials frequently encountered in work, home and community contexts. The proficiency score levels provide a useful indication of the range of tasks that an individual can be expected to perform with a high level of consistency. As well, the test scores can indicate the most appropriate materials to be used in a program. In addition, the test scores provide learners

with planning information and can assist in estimating individual learner growth and progress.

Test Development

The underlying principle driving the development of the simulation tasks was that demonstrated performance on any given task reflects the interaction among the structure of the stimulus material, the content represented and the nature of what the individual is asked to do with the material. Using Item Response Theory (IRT) in developing the TALS allowed for the estimation of the difficulty level of a particular exercise relative to the difficulties of all exercises in the pool, as well as the estimation of an individual's proficiency level in the assessed area. Four procedures were followed to establish the conversion tables that transform the raw score to the IRT estimated scale score: a) the establishment of IRT scales and estimation of national distribution from the 1985 Young Adult Literacy Survey b) the establishment of separate IRT scales with the TALS items c) the linking of the TALS scales to the Young Adult Literacy Survey scales via linking samples and d) the production of score conversion tables for the TALS.

Scale and Test Properties

Response probabilities. A particular benefit to the tests as a result of the application of an IRT model was the availability of item information that allows for the description of various proficiency levels over the range of each scale. This means it is possible to identify tasks that an individual scoring at a particular level can be expected with high probability to perform successfully, while a person scoring at the next lower score has a much lower probability of success.

Inter-scale relationships. Estimated correlations among the three literacy skills were .59 (Prose and Document), .60 (Prose and Quantitative) and .62 (Document and Quantitative).

Differential group performance. A deliberate attempt was made to select stimuli across a broad range of adult interests and contexts so that highly specialized background knowledge would not provide an advantage to any one group over another. As well, questions were developed to tap skills relevant to the types and levels of processing associated with the three literacy domains as reflected and practiced in everyday situations.

Basic English Skills Test

Centre for Applied Linguistics
1118 22nd Street, N.W.
Washington, DC 20037
Tel.: 202-429-9292

Test Content

The Basic English Skills Test (BEST) was developed to measure listening comprehension, speaking, reading and writing skills at a basic and intermediate level. It provides information on the attainment of basic functional language skills and consists of two sections: an oral interview section and a literacy skills section.

The oral interview section is made up of 49 items divided into personal identification, home environment, time, directions, shopping and money, signs, doctors office, accidents, employment and data forms. A reading passage and a writing task are also included. These two tasks serve as a screening device to identify examinees for whom the literacy skills section may be appropriate. All questions in this section of the test are a series of simulated real life listening, comprehension and speaking tasks. This part of the test is administered individually in a face to face interview and takes about fifteen minutes to complete.

The literacy skills section contains about fifty items categorized under personal background, calendar, food and clothing labels, rent cheque, envelopes, telephone directory, train schedule, signs, a reading passage and a writing question. This part of the test can be administered either individually or on a group basis and takes about an hour.

Description of Scores

Each item in the oral interview section of the BEST is designed to test one of four skills areas of listening comprehension, communication, fluency and reading and writing. Each of these four areas is also represented by a symbol on the test booklet. For each item of each skill area a range of scores is provided along with a summary of the oral interview scoring rules. This range includes values from 0-3 with a brief description and relevant examples of what is meant by the numerical weight. After scores are totalled using a scoring sheet, the test administrator derives an oral interview total score. Examinees who obtain a score of five or above can be given the literacy skills test.

A literacy skills scoring sheet is used for the second part of the test. Scoring is based on a scale of 1 or 0 points for the eleven parts of the test. The

test manual provides scoring instructions and relevant examples. Total scores for all questions are written on the scoring sheet. Using a competency based approach seven student performance levels are related to the different ranges of scores for both sections of the test.

Information from both sections of the test can be used for several purposes. Placement testing or entrance testing evaluates the extent and nature of the students language proficiency on entering an English language training course. Placing students according to BEST test scores can result in more or less homogeneous classes. The test can also be used to measure the progress of students in developing functional English language proficiency. It can determine progress made in a particular course of instruction provided the course is designed to teach survival and pre-employment language skills. A third use of the test can be for diagnostic purposes. The BEST is designed to reflect as closely as possible the actual language use situation with which students need to cope in independent daily living. An item by item review of a student performance on the test can provide important information on the student's ability or lack of ability to perform particular language use tasks. A final purpose of the test is for screening such as assessing the readiness for a particular course of study or training program.

Test Development

Content specifications for the test focused on topics identified as crucial to survival level competency in English. The total number of students who participated in the US field testing of forms B C and D was 987 for the oral interview section and 632 for the literacy skills section. Selection of the items for inclusion in the operational form of the test was based primarily on the statistical results of an item analysis. Level of difficulty and discrimination (r-biserial co-efficients) were examined for each test item. Results showed that very few field test items needed to be eliminated. Internal consistency reliability estimates for sub-scores of the three forms range from .77 - .83 and for total scores .96 - .97. These scores are quite high for a test composed mostly of free response items. Inter rater reliability or how different raters score the same examinee indicated that there was a high degree of consistency across raters.

The BEST demonstrated high face validity since its content is quite similar to real life language use tasks it aims to test. Some evidence is provided on construct validity when test scores were correlated with the pre-assigned group rating of the examinees' overall language proficiency. In terms of inter-subscale correlations, each subscale of the oral interview section of the test shows

substantial positive correlation with the total section score. As well there is a high correlation between reading and writing scores in all three forms.

Canadian Adult Achievement Test

The Psychological Corporation
Harcort Brace and company
555 Horner Avenue
Toronto, ON M8Z 4 X6
(1-800-387-7278)

Test Content

The Canadian Adult Achievement Test (CAAT) is designed to measure an adult's level of educational achievement. This battery includes subtest in Vocabulary, Reading Comprehension, Spelling, Number Operation, Problem Solving, Mechanical Reasoning, Language, Science, and Study Skills. There are four levels of CAAT: level A is for adults who have had one to three years of formal education; level B is for adults who have had four to six years of formal education; level C is for adults who have had at least seven to ten years of formal education and may or may not have finished high school and level D is for adults who have had eleven to twelve or more years of formal education and have probably finished high school. Not all levels include all subtests.

The CAAT assesses and describes learners' generic skills in such areas as phonics, reading school related text, identifying misspelled words, drawing inferences from different types of reading material, addition and subtraction or

consumer related problem solving abilities. To give the reader a sense of the content of these various subtests, a few are described below.

The vocabulary subtest is designed to assess the knowledge and understanding of words that are frequently encountered by adults in their work or daily activities. Thirty-four multiple choice items include words sampled from applied or general vocabulary, from vocabulary of the physical and natural sciences and from vocabulary of the social sciences. The problem solving subtest consists of twenty-one problems and measures an examinee's ability to determine an outcome, to record and retrieve information, and to use geometric concepts to measure. Using whole numbers, decimals, percents and fractions, the learner is requested to solve problems an adult is likely to encounter. The language subtest is taken from the Stanford Test of Academic Skills and is designed to measure a learner's functional knowledge and effective use of the English language. This subtest is made up of fifty-four items and is organized into four parts: reference skills, language sensitivity, conventions and paragraph arrangements.

Description of Scores

Two main kinds of scores can be obtained from the CAAT series: content referenced and norm referenced scores. An important score under the first category is the objective raw score. Average (mean) raw scores for all CAAT objectives at each level have been determined for the sample taking each level of the test as part of the CAAT research program. The performance of individual learners or groups of learners on each CAAT's objective can be evaluated in relation to those average scores for the purpose of identifying relative strengths and weaknesses.

Norm referenced scores are used to compare a learner's performance to that of an appropriate reference group. Grade equivalence and scale scores are types of norm referenced scores that can be used for this purpose when CAAT has been administered. Grade equivalence relate learners' scores to the typical performance of a number of groups of adults in specified grade tested in a given month of the school year. The CAAT grade equivalence ranges from K.0 to 12.9 with grade equivalence higher than 12.9 designated as post secondary. Despite the limitations, grade equivalence scores can provide an indication as to the approximate level at which adult instruction can be initiated.

CAAT scaled scores express performance on subtests and totals that extend across levels. For example, since the Reading Comprehension subtest appears in all three levels of CAAT, scaled scores for this subtest are comparable from level A to level B, to level C to level D. The continued nature of the scale and its equal interval property makes scale scores particularly suitable for studying change over an instructional period, regardless of the level of CAAT that was taken before and after instruction.

Test Development

Because CAAT is designed to assess the educational achievement of adults, it was considered necessary that adults, not school children be used in the standardization program. However, despite the fact that much work is being done at the present time concerning the measurement of adult achievement, there are still no suitable criteria for defining the population of adults across Canada for whom the CAAT would be appropriate. In the absence of such criteria, research was conducted with a number of adult groups whose characteristics would define the users of this type of instrument. The sample was drawn from lists of adult educational and vocational programs in communities and schools and from

mailing lists of provincial and regional and penal institutions. This research yielded information about the appropriateness of the items and objectives, the difficulty and sensitivity of the item for the adult group and the effectiveness of the options for all multiple choice questions.

Reliability. The reliability of a test is defined as the extent to which the test yields consistent results. The Kuder-Richardson Formula 20 was used to determine the reliability of CAAT's scores. Reliability estimates range from .84 - .95 for level A, B, and C.

Validity. The content validity of the test can be evaluated through the examination of the test content to determine whether it measures the fundamental skills that are necessary for an adult to function in today's society. Since the objectives were written to reflect the content of adult education programs throughout Canada, CAAT is valid for measuring these common objectives. However, individual programs must determine the validity of CAAT for measuring their own objectives. Evidence also indicated that inter-correlations of the CAAT are positively correlated with one another.

Measurement, Reading and Arithmetic

Ramsay Corporation
Boyce Station Offices
1050 Boyce Road
Pittsburg, PA 15241-3907
Tel.: 412-257-0732

Test Content

This criterion referenced test is made up of three basic skills in the areas of measurement, reading and arithmetic. The measurement of wholes, halves, quarters, eighths, and sixteenths comprise the first portion of the test intended for use where trainees must be able to measure accurately. Twenty items in a multiple format make up this section. The reading portion of the test consists of five paragraphs which are progressively more difficult in readability level. Each of the test passages is followed by eight multiple choice questions. The readability estimate range is from 8.0 for the instructions to 12.0 for the final passage. The average readability of the test is 10.6.

The job related passages are designed to assess ability to read job related information found in an industrial environment in the following areas: hydraulic fluids, bench tools, hydraulics, lubrication and solenoids. The arithmetic portion of the test was developed to evaluate the ability to perform basic computations of

addition, subtraction, multiplication and division of whole numbers, decimals and fractions. It also contains simple formulas and charts. Forty multiple choice items make up this part of the test. Each part of the test has a time limit. Measurement is timed at 15 minutes; reading takes 40 minutes to complete and the arithmetic subtest in 30 minutes.

Description of Scores

For this criterion referenced test, an absolute standard or criterion of performance is set and everyone's score is established in relation to that standard. Thus, 90 percent and above correct might be necessary to receive a grade of A, 80 to 89 percent correct for a B and so forth. Percentile equivalents for the three portions of test have been conducted on a sample of 1,441 metal workers.

Test Development

Reliability studies have been conducted on other tests from the Ramsey Corporation such as the Technician Electrical Test, the Technician Mechanical Test, the Building Maintenance Test, the Electronics Test and the Electrical Maintenance Trainee Test but not this newly developed one. In terms of the basic

skills test validation procedures, seven steps have been taken. First, the client provides a company job description which is then placed into the Ramsay Corporation format of job summary, work performed and consequences of errors. Ten subject experts add to the description to ensure that it is up to date and reflects the job activities. Then the ten subject experts rate the job with respect to importance of job activities and percent of time spent on job activities. The inter rater agreement is then evaluated. The third step is the evaluation of the readability levels, mathematical skills and other requirements of the job and the training program. The testing company then evaluates the job analysis information and determines the knowledge, skills and ability and the methods for measurement. The fifth step entails editing, composing and printing the test. They are brought to a group of at least four subject experts for review and matching with knowledge, skills and ability. The Angoff procedure to set cutting scores and review the final test are used. This means that 5 to 10 subject experts review the test item by item, select a correct answer, are told the keyed answer and are asked what percent of qualified candidates would pass each item. Difficulty and point biserial discrimination indices have been calculated. The last step includes the writing of a content validation report in the format suggested by the Uniform Guideline on Employee Selection Procedures. This also means that

the company reviews a draft and the final report detailing the activities undertaken is bound and provided to the client.

Wide Range Achievement Test 3

Jastak Associates
P.O. Box 3410
Wilmington, DE
19804-0250
Tel.: 1-800-221-9728

Test Content

The Wide Range Achievement Test 3 (WRAT3) is designed to measure the codes which are needed to learn the basic skills of reading, spelling and arithmetic. This recent addition of the test can be used with adults ranging in ages from 17 to 75. Each of the two alternate test forms provide three subtests. The first subtest is reading which consists of a letter reading section and a word reading section. In the letter reading section the respondent is asked to name 15 letters of the alphabet and in the word reading section he is asked to pronounce 42 words. Both parts of the reading subtest are to be administered individually. For adults, the word reading section is administered first. It is a word recognition test and not a test of speech or diction. If a more comprehensive reading test is required both test forms can be administered. One point is given for each correct letter and word. Raw scores are transferred to a section on the test form.

For the second subtest spelling, respondents write letters of the alphabet and words to dictation. There are 55 items on both forms of the spelling test. These consist of 15 items on the letter writing section and 40 items on the word spelling section. For adults, the word spelling section is administered first. A pronunciation guide is provided for the test administrator with the list of spelling words on plastic cards. One point is given for each of the correctly written letters and spelled words. Total spelling raw scores are then transferred to a section on the test form.

The arithmetic subtest consists of counting, reading number symbols, solving oral problems and performing written computations. It is also broken down into an oral arithmetic section with 15 items and a written arithmetic section with 40 items. Adults begin with the written arithmetic section. Each item is counted as one point and the total arithmetic raw score is transferred to a section on the test form. Each form of the WRAT3 takes 15 to 30 minutes to complete.

Description of Scores

There are several types of scores used to report the WRAT3 results. Four kinds of scores will be highlighted here. The raw scores are the number counts of the correct responses. They cannot give a precise measure of how much better one performance is over another. The absolute scores of an individual or group of individuals give measurement of each variable across the whole continuum of that variable without regard to grade or age. These scores are most suitable for pre- and post-testing. Standard scores are the types of scores used for comparisons within individuals and between them. Not only is it possible to tell that there is a difference between the two scores, it is appropriate to describe the amount of difference between such scores. Examples of the ratings of standard scores are: high average (110-119), average (90-109), low average (80-89). Grade scores are slightly higher level scores than raw scores. For the WRAT3, grade scores are only rough guides to grade achievement because there is as much variance within a grade as there is between grades. These scores should be used with the information from standard scores or absolute scores in describing the test results.

Test Development

The WRAT3 is an age-normed referenced test of the basic academic coding skills of word recognition, spelling and arithmetic computation. Re-standardization of this edition of the test began in the early 90s as a result of the Rasch analysis of item difficulties. It was found that there were many duplications of items in bold forms of the WRAT-R. The WRAT3 was given to a norm sample in the United States. An adult sample was selected by using a national stratified sample design controlled for age, regional residence, gender, ethnicity and socio-economic level. The sample size for all age groups(5-75) was 4,834.

Reliability. When a skill is measured there should be a reasonable expectation that another measurement of the same skill will give similar results. The extent that an instrument can make consistent measurement of a trait is described by the reliability of the test. Three of the many measures of reliability are reported here. The median test coefficient alphas range from .85 - .95 over the different test forms. This is considered desirable. Correlation coefficients for the raw scores of the three subtests are also very high — .98, .98 and .98 respectively. Using the test-retest method corrected stability coefficients range from .91 to .98.

Content validity. The Rasch statistics of item separation was used to indicate how well the items selected were representative of the three domain areas: written decoding, written encoding and arithmetic computation. For each of the subtests the highest item separation score possible 1.00 was found which provides strong evidence that there is content validity on each of the three variables.

Construct validity. Evidence regarding the developmental nature of the WRAT3 shows a steady increase in mean scores across the norm age groups until the 45-54 year old age groups. From here the scores begin to fall. Using the data from the total norm sample, there is a significant correlation between each WRAT3 test raw score and the age and month of the test taker. Evidence also indicated that inter-correlations of the WRAT3 are positively correlated with one another. The range of correlation for the reading-spelling comparison is .81 - .91; reading-arithmetic .54 - .78; and spelling-arithmetic .58 - .82. The inter-correlations show a moderate to high positive relationship among the variables.

Metric Skills Assessment

Lakeshore Technical College
Gary Heeler
1290 North Avenue
Cleveland WI 53015
Tel.: 414-458-4183

Test Content

The Metric Skills Assessment is a competency-based test designed to measure a predeveloped series of metric learning experiences. The test which consists of 35 items is divided into two sections. The first section covers areas of metric origin, basic units of measure, definitions, and numerical conversions. The second section is a series of questions asking the test taker to record measurements provided by different job-related instruments such as the micrometer and the vernier caliper. The test can be administered individually or to groups of adult learners and takes about 20 minutes. The test also can be used as a pretest/post-test training tool.

Description of Scores

The Metric Skills Assessment is a type of criterion-referenced tool which means that learner gain is interpreted in terms of how many units of instruction

are mastered at the prescribed criterion level. The test can be administered as a training program pretest and depending on the criterion used by the instructor can direct the learner into the appropriate part of the curriculum. As a post-test, the predetermined criterion is usually 90% correct or 80% correct.

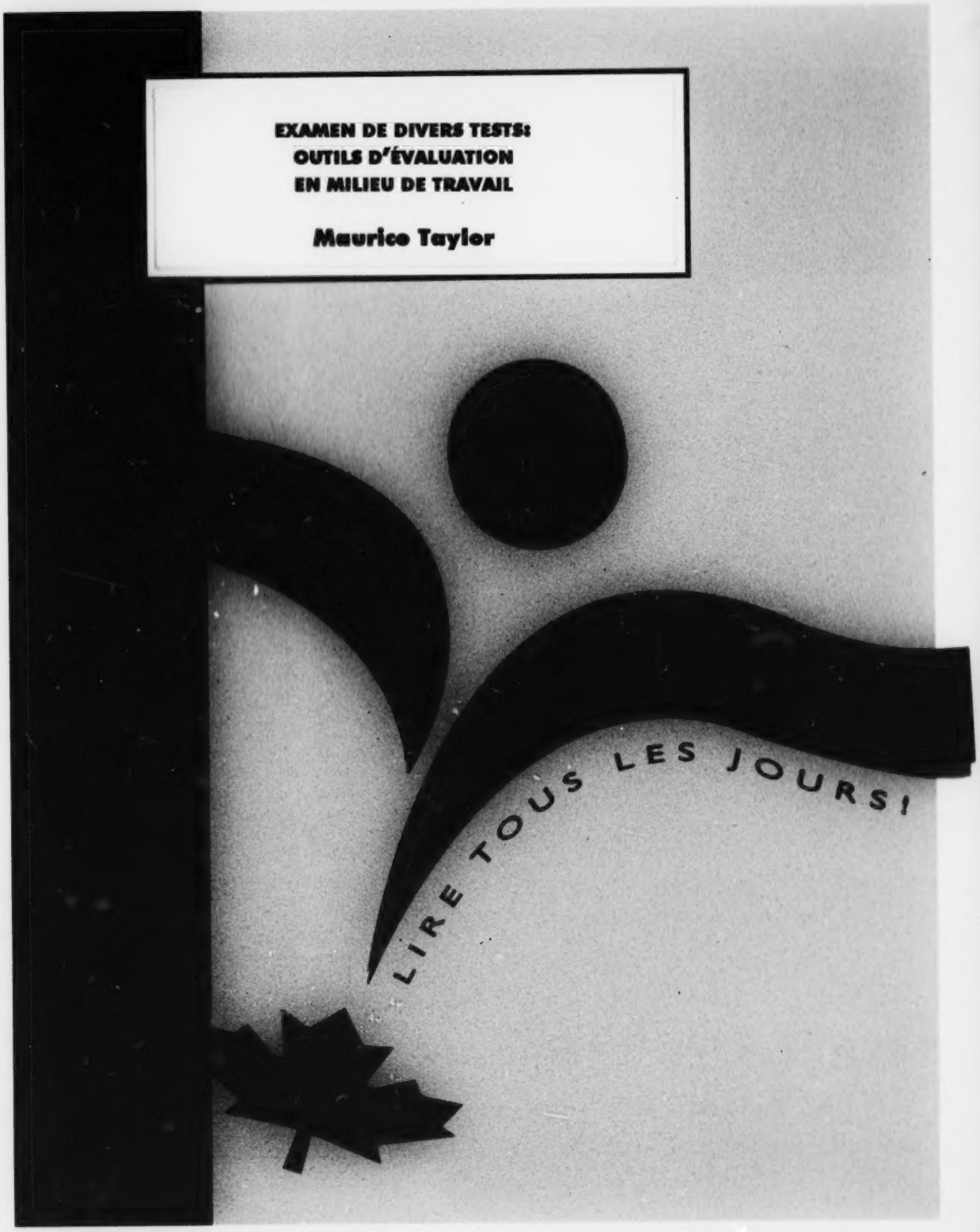
Test Development

The test developer first identified what important competencies should be learned through a panel of mathematical experts. Then various commercially available curriculum materials with a variety of learning exercises were identified that teach each of the competencies. In other words, the competencies were first identified and a curriculum was designed to help the learner achieve those competencies. The curriculum is divided into six key areas. In the first module the everyday units of the metric system are introduced with a review of decimal operations. Traditional and metric measurement and the use of metric scales and tapes to measure items are also taught. The focus of the second module is conversion between metric and English units of measure while the third area centers on metric thread pitch and hardware identification. In the fourth module,

precision measuring equipment use and identification is taught while the fifth area focuses on the use of verniers. A review of the metric system of measurement and skills reassessment concludes the program.

EXAMEN DE DIVERS TESTS: OUTILS D'ÉVALUATION EN MILIEU DE TRAVAIL

Maurice Taylor



Le Secrétariat national à l'alphabétisation

National Literacy Secretariat

Développement des ressources humaines Canada

Human Resources Development Canada

Canada

BEST COPY AVAILABLE

Canada

EXAMEN DE DIVERS TESTS:
Outils d'évaluation en milieu de travail

Maurice Taylor

Partnerships in Learning
Avril 1997

Développement des ressources humaines Canada
Human Resources Development Canada
Le Secrétariat national à l'alphabétisation
National Literacy Secretariat

Pour commander des exemplaires du présent rappport, communiquez avec:

Le Secrétariat national à l'alphabétisation
Développement des ressources humaines Canada
Ottawa (Ontario) K1A 1K5
Téléphone: (819) 953-5280
Télécopieur: (819) 953-8076
ATME: (819) 953-2338
Courrier élect: nls@fox.nstn.ca

Table des matières

Partie 1 Dimensions du contrôle des connaissances

L'évaluation en situation d'apprentissage..... 3

Quelles sont les compétences essentielles dans le monde d'aujourd'hui?.....6

Types de test.....11

Partie 2 Examen technique détaillé de divers tests

Tests of Adult Basic Education (TABE) Work-Related Foundation Skills.....14

Tests of Applied Literacy Skills 20

Basic English Skills Test 27

Canadian Adult Achievement Test 33

Measurement, Reading and Arithmetic.....39

Wide Range Achievement Test 3 43

Metric Skills Assessment.....49

INTRODUCTION

L'évaluation des participants aux programmes d'alphabétisation en milieu de travail est toujours l'objet de vifs débats dans le domaine de l'éducation des adultes. C'est une question qui peut être abordée de divers points de vue , sous l'angle tantôt des composantes essentielles de tout test normalisé tantôt des moyens d'obtenir l'information nécessaire à la prise de décisions. Dans le document qui suit, nous nous intéressons à un aspect de l'évaluation, à savoir le contrôle des connaissances. Nous procédons à un examen technique détaillé de sept outils d'évaluation employés dans le domaine de l'éducation en milieu de travail. Nous examinons chaque instrument au regard du contenu, des scores et du processus d'élaboration.

Pour bien situer le sujet dans son contexte, nous commençons par examiner certaines des dimensions du contrôle des connaissances. Cet examen préalable pourra servir de point de départ pour aider les praticiens à choisir les tests à utiliser quand il faut évaluer les participants à un programme d'éducation en milieu de travail. Le but du document qui suit est de montrer, non pas que certains tests sont supérieurs à d'autres, mais que les formateurs en milieu de travail doivent trouver le test qui convient le mieux aux besoins de leurs apprenants et de leurs programmes respectifs.

PARTIE 1

DIMENSIONS DU CONTRÔLE DES CONNAISSANCES

L'évaluation en situation d'apprentissage

L'évaluation dans le domaine de l'éducation en milieu de travail peut être considérée comme une carte routière qui aide les gens à arriver à une destination en particulier. La carte ou le plan d'apprentissage guide le stagiaire vers une destination qui correspond aux buts et objectifs du programme. Les gens se servent souvent de cartes routières pour déterminer la route à prendre pour arriver à leur destination ; certains veulent la route la plus rapide, tandis que d'autres préfèrent une route plus pittoresque qui comprend des points d'intérêt. Il en est de même pour l'évaluation, qui devrait aider l'apprenant à avancer vers ses objectifs et lui indiquer combien de temps il lui faudra pour les atteindre. Dans beaucoup de programmes d'éducation offerts actuellement en milieu de travail, l'évaluation est utilisée pour répartir les apprenants selon leur niveau, pour leur prescrire des plans d'apprentissage propres à leurs besoins, pour contrôler leurs progrès et pour

vérifier la réalisation des objectifs. Beaucoup de praticiens et d'universitaires estiment que l'évaluation bien faite est un élément essentiel du plan d'apprentissage au même titre que le programme d'études et l'enseignement. Le plan d'apprentissage est constitué de diverses stratégies et techniques d'enseignement, d'évaluation et de renforcement de l'apprentissage. Aucune méthode d'enseignement ne peut être efficace pour tous les apprenants dans toutes les circonstances ; de même, aucun instrument ni procédé d'évaluation ne permet de fournir de l'information de qualité qui réponde à tous les besoins.

De manière générale, l'évaluation dans le domaine plus vaste de l'alphabétisation des adultes se fait selon trois grandes approches, qui peuvent être utiles pour mieux comprendre l'évaluation dans le domaine de l'alphabétisation en milieu de travail. Chaque approche repose sur certaines hypothèses et comporte certaines limites. Selon l'approche traditionnelle, l'alphabétisation fonctionnelle vise l'obtention aux tests de rendement d'un score correspondant à un niveau scolaire en particulier. Ces tests permettent d'évaluer le pourcentage de divers groupes dont le rendement est égal ou supérieur à un niveau scolaire particulier en lecture. Ceux dont le rendement est égal ou supérieur au niveau en question sont

présumés avoir les compétences voulues pour accomplir des tâches ou se servir de documents correspondant à un niveau de difficulté comparable.

La deuxième approche se fonde sur l'acquisition de compétences. Des tests comportant des tâches élaborées à partir de matériel de type non scolaire sont administrés à des échantillons nationaux. L'analyse des résultats obtenus permet d'évaluer la proportion de la population adulte qui devrait pouvoir accomplir avec succès les tâches en question et de déterminer dans quelle mesure diverses caractéristiques, comme le niveau d'instruction, l'origine raciale, l'origine ethnique, le sexe et le niveau de revenu, sont liées aux taux estimatifs de réussite.

Selon l'approche profil, l'alphabétisation est considérée, non pas simplement comme l'acquisition d'un ensemble de compétences isolées liées à la lecture et à l'écriture, mais davantage comme l'application de ces compétences à des fins précises et dans des contextes précis. Étant donné la diversité des contextes d'application, au foyer, au travail et en société, l'alphabétisation est perçue, non pas comme une norme unique, mais comme un ensemble de pratiques utilisées dans divers contextes. Cette approche comprend trois aspects distincts de l'alphabétisation : les textes, les documents et les nombres.

Le lecteur constatera que chacune de ces trois façons d'aborder l'évaluation se trouvent représentées dans la partie examen technique détaillé de divers tests.

Quelles sont les compétences essentielles dans le monde d'aujourd'hui ?

Les divers types de compétences nécessaires dans le monde du travail sont une autre dimension de l'évaluation et du contrôle des connaissances. Pour choisir un instrument d'évaluation, il faut notamment se demander si les compétences qu'il évalue sont du même type que les compétences devant être appliquées dans l'exercice des fonctions actuelles ou futures. C'est là un domaine d'enquête qui suscite l'intérêt depuis déjà un certain temps. Parmi les travaux de recherche en cours, il convient de signaler le Projet de recherche sur les compétences essentielles de Développement des ressources humaines Canada, qui vise à déterminer ce que font les travailleurs de divers secteurs. Les chercheurs ont interviewé quelque 3 000 travailleurs d'une multitude de secteurs au Canada afin de répertorier les compétences essentielles effectivement nécessaires pour les occupations de premier échelon exigeant des compétences minimes. L'étude

devait se solder par un profil décrivant les compétences essentielles pour chacune de ces occupations dans le contexte canadien. Les profils ainsi obtenus comprennent aussi bien les compétences effectivement utilisées au travail telles qu'elles sont décrites par les travailleurs en poste que les compétences de base traditionnelles, soit la capacité de raisonnement, la capacité de travailler en équipe et les compétences nécessaires à la formation permanente. Quand on a une idée générale de ces différentes catégories de compétences, on peut mieux choisir le type d'instrument d'évaluation à utiliser. Il convient de signaler ici qu'aucun des sept tests examinés dans la suite du document n'est destiné à mesurer toutes les catégories de compétences décrites dans les paragraphes qui suivent.

La lecture de textes, la recherche documentaire, la rédaction, le calcul et la communication oraie sont au nombre des compétences de base traditionnelles. Voici des exemples de compétences qu'on retrouve dans certaines de ces catégories. Dans la catégorie lecture de textes, certaines tâches typiques ou plus complexes sont évaluées selon une échelle qui va de la lecture de textes relativement courts jusqu'à l'interprétation de textes denses et complexes qui comporte un niveau de difficulté plus grand, en passant par la recherche d'un

élément d'information en particulier ou la compréhension d'instructions écrites simples.

La catégorie recherche documentaire comprend des tâches typiques ou plus complexes en lecture. Ces tâches sont évaluées selon le niveau de difficulté et vont de la lecture d'affiches, d'étiquettes et d'autres textes semblables où il suffit de pouvoir retrouver des mots clés jusqu'aux déductions fort abstraites faites à partir de textes et à l'utilisation de connaissances spécialisées. Dans la catégorie rédaction, on retrouve des tâches typiques ou plus complexes selon le cas qui sont aussi évaluées, avec au bas de l'échelle les compétences comme la rédaction de textes très courts à insérer dans un formulaire et au haut les tâches plus complexes comme la rédaction de textes sur une vaste gamme de sujets en vue persuader quelqu'un ou de présenter une analyse, une comparaison, une évaluation ou une critique.

Dans la catégorie calcul, les tâches sont réparties par domaine d'application, selon qu'il s'agit d'opérations financières simples, de l'établissement d'horaires ou de budgets, de mesures ou d'analyse de données. La capacité de calcul englobe aussi des compétences plus précises ayant trait aux entiers, aux fractions, aux proportions, à la conversion de mesures et à la géométrie. La capacité de communication orale est évaluée selon le type de communication et le but visé,

comme d'écouter, de parler, de traiter avec les camarades de travail, de traiter avec les superviseurs, de traiter avec les clients, de participer à des discussions de groupe, de présenter de l'information à des groupes nombreux et d'animer des discussions structurées.

La deuxième grande catégorie de compétences concerne la capacité de raisonnement, qui comporte diverses sous-catégories. La première sous-catégorie est celle des compétences liées à la résolution de problèmes, qui sont elles aussi évaluées selon divers facteurs, comme la complexité du problème, la difficulté de le définir, la difficulté de définir les étapes permettant d'en arriver à une solution et la difficulté d'évaluer la situation. Vient ensuite la prise de décisions qui est liée à la résolution de problèmes. Encore là, les tâches typiques ou plus complexes sont cotées selon une échelle qui tient compte de facteurs comme la présence d'un ensemble de procédures ou d'un arbre de décision dont on peut s'inspirer, la conséquence des erreurs et le degré de jugement nécessaire pour prendre une décision acceptable. Toujours sous la rubrique des compétences liées à la capacité de raisonnement, on trouve la sous-catégorie des compétences permettant de planifier et d'organiser son travail. Ces compétences sont cotées selon une échelle qui tient compte de facteurs comme la diversité des fonctions à

accomplir et la mesure dans laquelle le plan de travail de l'employé doit s'intégrer aux plans de travail d'autres employés. Viennent ensuite les compétences liées à la mémoire. Ces compétences sont évaluées selon le type de mémoire auxquelles elles font appel : se rappeler une certaine information pendant un court laps de temps, se rappeler des événements précis et se rappeler des informations propres à son travail pour s'en servir à long terme.

Dans la troisième grande catégorie, la capacité à travailler avec d'autres, les compétences sont réparties selon le contexte dans lequel le travail se fait, comme de travailler seul, de façon indépendante, avec un partenaire ou un assistant ou comme membre d'une équipe, et selon qu'il s'agit de jouer un rôle de supervision ou de leadership, comme d'orienter les nouveaux employés, d'attribuer des tâches à d'autres travailleurs, de choisir des entrepreneurs ou des fournisseurs ou de participer à des discussions structurées sur la façon dont le travail doit se faire. Les compétences informatiques constituent la catégorie suivante. L'utilisation d'ordinateurs est évaluée selon une échelle de un à cinq points, le plus bas niveau étant celui qui consiste à se servir d'un ordinateur sans avoir besoin de connaître le logiciel requis et le niveau le plus élevé étant celui qui consiste à évaluer les besoins et à concevoir des systèmes informatiques.

En résumé donc, les compétences essentielles sont celles qu'on retrouve dans presque tous les types d'occupation, même si la forme ou le niveau de complexité peut varier selon le type d'occupation. Ces compétences sont essentielles à l'acquisition des compétences plus précises qui permettent d'accomplir les tâches propres à un type de travail en particulier.

Types de tests

Beaucoup de praticiens trouvent difficile de prendre des décisions judicieuses quant aux outils d'évaluation à utiliser parce qu'ils ont du mal à s'y retrouver dans la terminologie technique des tests. Bien que le sujet ait déjà été abordé dans d'autre documents, la section qui suit décrit certains des termes du vocabulaire des tests. Normalisation, normes, norme absolue et tests axés sur les compétences sont autant de termes que le lecteur rencontrera dans la partie examen technique détaillé de divers tests. Test normalisé s'entend d'un test qui est administré dans des conditions normalisées et qui permet de faire des déductions au sujet de la capacité de l'apprenant à partir d'un échantillon de performances étalonnées. Ce qui distingue le test normalisé du test informel, c'est

qu'il est administré dans des conditions uniformes. Ainsi, le test de rendement normalisé comporte les mêmes textes et les apprenants doivent suivre les mêmes procédures, répondre aux mêmes types de questions et respecter les mêmes limites de temps.

Il existe deux types de tests normalisés : le test normatif et le test critériel. Beaucoup de tests normalisés sont conçus pour permettre d'interpréter le score de l'apprenant par rapport ou par référence aux scores obtenus par d'autres sujets qui ont fait le test. C'est ce qu'on appelle le test normatif. Le score qu'obtient le sujet est interprété en le comparant aux performances étalonnées du groupe de référence.

L'autre type de test normalisé est le test critériel. L'utilisation de ce type de test va de pair avec l'élaboration de programmes d'enseignement individualisés où chacun apprend à son rythme. Selon cette approche, le champ des connaissances ou des compétences à acquérir est soigneusement défini. Le score de chaque apprenant est déterminé par rapport à une norme ou un critère de performance absolu. Les programmes de ce genre exigent d'établir des objectifs d'apprentissage précis et de prévoir des unités d'enseignement, souvent appelées

modules, pour enseigner les divers sous-ensembles de connaissances et de compétences définis dans les objectifs d'apprentissage.

L'évaluation peut aussi se faire à partir de tests axés sur les compétences. La première étape de l'élaboration d'un test de ce genre consiste à établir les objectifs clés à atteindre ou les principales compétences à acquérir. On arrête ensuite les éléments du test de manière à déterminer si l'apprenant possède déjà les compétences voulues ou s'il a besoin d'un programme d'enseignement pour acquérir certaines compétences. Enfin, on dresse la liste des divers matériels pédagogiques offerts sur le marché et qui comportent une gamme variée d'exercices visant l'acquisition de chacune des compétences de façon que les enseignants puissent choisir ce qui convient à leurs apprenants. Selon cette approche, l'enseignement se fait en quelque sorte en fonction du test, même si le contenu comme tel du test ne figure pas nécessairement au programme d'études afin d'éviter que l'enseignement ne soit directement lié aux divers éléments du test. Les programmes axés sur les compétences exigent donc qu'on définisse d'abord les compétences à acquérir et qu'on structure le programme d'études afin d'aider l'apprenant à acquérir ces compétences.

PARTIE 2

TESTS OF ADULT BASIC EDUCATION (TABE)
(tests d'éducation de base pour adultes)

Work-Related Foundation Skills
CTB MacMillen-MacGraw Hill
20, chemin Ranch
Monterey (Californie)
93940-5703
Tél. : 1-800-538-9547

Contenu

La Work-Related Foundation Skills (WRFS) est une collection de tests normatifs visant à déterminer dans quelle mesure le sujet possède les compétences de base en lecture, en calcul et en activités langagières qu'on retrouve dans la majorité des programmes d'éducation de base pour adultes. À l'aide d'un contenu exprimé dans un langage judicieusement choisi pour représenter le contexte du milieu de travail, ces tests déterminent dans quelle mesure le sujet possède les compétences voulues pour fonctionner au travail et en société et quels sont ses besoins à cet égard.

La WRFS comprend des tests dans quatre domaines : santé, métiers-professions techniques, commerce-bureau et général. Les questions, les graphiques et les passages qui composent chaque test sont présentés dans un contexte apparenté à ces différents milieux de travail. Les quatre tests de la collection WRFS mesurent des compétences comparables. Comme leur contenu est comparable, les quatre tests peuvent être utilisés efficacement soit avant soit après la formation.

Chacun des quatre tests de la WRFS comprend les sous-tests suivants : lecture, calcul arithmétique, mathématiques appliquées et activités langagières. Chaque sous-test dure une trentaine de minutes. Le sous-test de lecture porte sur la compréhension du processus de lecture ainsi que sur la capacité à trouver le sens d'un texte. À l'aide de passages écrits, de graphiques et de références, le sous-test évalue la capacité du sujet à reconnaître des mots dans un contexte de travail, à se rappeler des informations et à retrouver les idées principales. Les questions visent à mesurer la capacité de lecture fonctionnelle de diagrammes, de tableaux, de formulaires et d'horaires.

Le sous-test de calcul arithmétique vise à déterminer dans quelle mesure le sujet possède les compétences de base voulues pour suivre avec succès un

programme d'enseignement en mathématiques, dans quelle mesure il comprend les différentes opérations arithmétiques, addition, soustraction, multiplication et division, avec des entiers, des décimales et des fractions. Il vise aussi à mesurer sa compréhension des opérations financières simples, des entiers et des pourcentages. Le sous-test de mathématiques appliquées porte principalement sur la formation de concepts, le raisonnement et la résolution de problèmes en contexte. Il mesure la capacité du sujet à appliquer des concepts mathématiques liés à la rémunération, au calcul en contexte, à la théorie des nombres, aux mesures, à l'interprétation de données, à la géométrie et à la pré-algèbre. Il évalue aussi la capacité de raisonnement nécessaire à la résolution de problèmes pratiques.

Le sous-test de langue mesure les compétences du sujet en ce qui concerne les règles de l'usage, la formation de phrases, le développement de paragraphes et les conventions d'écriture comme l'utilisation des majuscules et la ponctuation. Le sous-test porte sur les compétences essentielles pour être considéré comme pleinement alphabétisé et met l'accent sur le lien entre langage, lecture et écriture. Il s'intéresse aussi à l'utilisation correcte du langage dans le milieu de travail. Le test présente une multitude de situations de travail authentiques en guise de stimuli.

Scores

La collection WRFS fournit des scores normatifs sous forme d'échelles de scores, d'équivalents de niveau scolaire et de rangs centiles. Il fournit aussi des informations critérielles relatives aux objectifs incorporés aux tests. Les sous-tests et les objectifs de la collection WRFS ont été établis à partir d'informations tirées des programmes d'éducation des adultes de divers États et après examen des résultats et des profils de compétences de programmes d'éducation des adultes à l'échelle des États-Unis. L'échantillon devant servir d'étalon pour les tests de la WRTS comprenait environ 1 700 apprenants adultes sélectionnés à l'échelle du pays en 1994. Les tests TABE de la collection WRTS de même que les formulaires 7 et 8 ont été administrés à chacun des participants.

Processus d'élaboration

Les tests de la collection WRFS ont été élaborés selon un processus typique comprenant les sept étapes suivantes : conception de l'évaluation, rédaction, examen des items, essai, analyse des items, choix du test et élaboration des normes.

Fiabilité. Les tests de la WRFS obtiennent un coefficient Kuder-Richardson formule 20 variant entre 0,82 et 0,91. Le coefficient KR 20 mesure la validité des scores. Plus il est élevé, plus les différents sous-test témoignent d'une cohérence interne élevée. Par cohérence interne, il faut entendre que chaque sous-test mesure dans son ensemble une seule compétence, comme la capacité de lecture, plutôt qu'une combinaison de plusieurs compétences.

Validité. La validité du contenu des tests de rendement tient à la concordance entre le contenu du test et le contenu du programme d'enseignement. Pour assurer cette concordance, les responsables de l'élaboration des tests ont examiné une multitude de programmes d'études et ont rencontré des spécialistes en éducation afin de déterminer les objectifs de rendement communs aux divers programmes existants de même que les connaissances et les compétences sur lesquelles on insiste dans ces programmes. L'information ainsi obtenue les a guidés dans toutes les étapes de la conception et de l'élaboration de cet outil d'évaluation des programmes d'éducation en milieu de travail.

Validité conceptuelle. La validité conceptuelle, qui englobe l'interprétation des scores et les déductions qui peuvent être faites à partir de ces scores, est un élément essentiel du processus de validation des outils d'évaluation utilisés en milieu de travail. Il existe d'abondantes preuves de la validité conceptuelle, preuves qui ont trait tant au contenu qu'aux critères. Ainsi, les outils d'évaluation de l'éducation en milieu de travail sont exhaustifs dans la mesure où ils comprennent des éléments qui reflètent les objectifs d'enseignement essentiels. En outre, les corrélations entre ces outils d'évaluation et entre les outils et les autres instruments de la famille des instruments d'évaluation TABE témoignent d'une validité convergente et divergente. Autrement dit, les tests qui sont conçus pour mesurer des compétences semblables ont une corrélation plus élevée que les tests conçus pour mesurer des compétences tout à fait différentes. Ainsi, le sous-test de langue de la WRFS est en étroite corrélation avec le sous-test de lecture, comme on pouvait s'y attendre, tandis que la corrélation avec les sous-tests de mathématiques de la WRFS est moins élevée.

TESTS OF APPLIED LITERACY SKILLS
(tests mesurant le degré d'alphabétisation fonctionnelle)

Educational Testing Service
Language Learning and Assessment
Chemin Rosedale
Arrêt postal 05-9
Princeton (New Jersey) 08541
Tél.: 1-800-551-1230

Contenu

Les tests du degré d'alphabétisation fonctionnelle (TALS) mesurent trois aspects distincts et importants du degré d'alphabétisation : la lecture de textes, la recherche documentaire et le calcul. Il existe deux formulaires pour chacun des trois sous-tests. Les tâches relatives à la lecture de textes exigent les connaissances et les compétences nécessaires pour comprendre et utiliser l'information contenue dans divers textes, notamment des éditoriaux, des articles de journaux et des poèmes. Les deux formulaires pour lé test de lecture de textes comprennent chacun 48 tâches. Les textes utilisés sont surtout de nature descriptive. Ils sont reproduits intégralement suivant la mise en pages et la typographie de la source initiale. Le sous-test évalue trois aspects de la lecture de textes : trouver de l'information, l'intégrer et la produire. Les tâches liées à

chacun des ces trois aspects comportent des niveaux de difficulté divers, tout comme celles qui sont incluses dans les trois sous-tests. Une période de 20 minutes est accordée pour chacune des deux sections de chaque formulaire.

Les tâches liées à la recherche documentaire englobent les connaissances et les compétences nécessaires pour trouver et utiliser l'information contenue dans des demandes d'emploi, des formulaires de paye, des horaires de transport, des cartes routières, des tableaux, des index, etcetera. Les deux formulaires pour le test de recherche documentaire comprennent chacun 52 tâches.
L'accomplissement de beaucoup de ces tâches est nécessaire pour répondre aux exigences d'un emploi et gérer un ménage. Le test mesure la maîtrise des techniques nécessaires pour trouver de l'information dans divers documents complexes, comme des graphiques et des tableaux, et pour transférer cette information d'un document à un autre. Trois types de questions ou d'instructions, amenant le sujet à trouver de l'information, à la parcourir et à l'intégrer, sont utilisés pour évaluer les compétences.

Les tâches liées au calcul englobent les connaissances et les compétences nécessaires pour faire des opérations arithmétiques, soit isolément soit en ordre séquentiel, à partir de chiffres figurant dans des documents : par exemple, établir

le solde d'un compte de chèque, fixer un itinéraire de voyage, remplir un formulaire de commande ou déterminer le montant de l'intérêt sur un prêt ou un placement. Pour le test de calcul, on peut choisir entre deux formulaires comportant chacun 46 tâches. Le sujet doit faire des additions, des soustractions, des multiplications et des divisions à partir de chiffres compris dans des documents qu'on retrouve dans la vie de tous les jours. Le degré de difficulté des tâches est fonction des opérations de calcul exigées, de la mesure dans laquelle les chiffres figurant dans les documents s'imposent à l'esprit et des déductions qui doivent être faites, le cas échéant, pour déterminer le type d'opérations nécessaires.

Scores

La première étape de l'interprétation des résultats du test consiste à calculer le score brut. Le score brut correspond au nombre total de bonnes réponses dans chaque section du formulaire, les réponses étant notées à partir du guide inclus dans le manuel qui accompagne le test. La deuxième étape consiste à convertir les

scores bruts en scores de performance à partir des tableaux qui se trouvent dans le manuel correspondant à chacun des formulaires des trois tests. Pour déterminer le score de performance du sujet, l'évaluateur choisit le tableau qui correspond au formulaire afin d'y trouver le nombre de bonnes réponses pour chacune des deux sections du test. Le chiffre marquant dans le tableau l'intersection des rangées correspondant à la section visée est le score de performance. Ce score donne une estimation de la compétence qu'a démontrée le sujet dans chacun des domaines de compétences à évaluer.

Une fois qu'il a en main le score de performance, l'évaluateur se sert du cadre d'interprétation inclus dans le manuel, qui permet de repérer sur chacune des échelles d'alphabétisation le niveau de difficulté correspondant aux différentes tâches. Les échelles sont conçues pour mesurer toute la gamme des degrés de difficulté de 0 à 500, mais la majorité des tâches se situent entre 200 et 400. Pour comprendre ce que chaque test vise à mesurer, les tâches sont interprétées à la lumière des interactions entre les stimuli et les questions ou instructions qui permettent de les situer sur chacune des échelles de difficulté. Le cadre d'interprétation du test de lecture de textes porte sur la capacité à retrouver de l'information dans un texte, à l'intégrer et à produire de l'information. De même,

le cadre d'interprétation du test de recherche documentaire porte sur la capacité à retrouver de l'information dans divers documents, à parcourir l'information et à l'intégrer. Les conditions et les niveaux de difficulté du test de calcul mettent l'accent sur les opérations d'addition, de soustraction, de multiplication et de division et sur des combinaisons de ces différentes opérations.

Pour l'apprenant, les tests mesurant le degré d'alphabétisation fonctionnelle peuvent être utiles à maints égards. Ils lui fournissent une estimation de sa capacité démontrée à appliquer diverses compétences en lecture, en écriture et en calcul à certains documents qui font partie de la vie de tous les jours, au travail, au foyer et en société. Les scores de performance sont une mesure efficace de la gamme de tâches que le sujet devrait pouvoir systématiquement accomplir avec succès. Les scores obtenus peuvent aussi indiquer quel serait le matériel pédagogique qui conviendrait le mieux aux besoins. En outre, les scores fournissent à l'apprenant des informations utiles pour planifier son programme d'études et lui permettent d'évaluer son évolution et ses progrès.

Processus d'élaboration

Les tâches de simulation ont été élaborées à partir du principe que la capacité démontrée à accomplir une tâche donnée reflète l'interaction entre la structure de l'élément stimulus, le contenu représenté et ce que le sujet doit faire de l'élément en question. Les tests ayant été élaborés à partir de la théorie de la réponse d'item (IRT), il est possible d'évaluer le niveau de difficulté d'un exercice donné relativement au degré de difficulté de tous les exercices visés et d'estimer aussi le niveau de rendement du sujet dans le domaine évalué. La création des tableaux de conversion qui permettent d'obtenir à partir du score brut le score gradué estimatif sur l'échelle IRT comportait quatre étapes : a) l'établissement d'échelles IRT et l'estimation de la répartition nationale des personnes ayant participé à l'enquête Young Adult Literacy Survey (enquête sur le degré d'alphabétisation des jeunes adultes) de 1985 ; b) l'établissement d'échelles IRT distinctes à partir des éléments des tests TALS ; c) l'établissement de liens entre les échelles TALS et les échelles Young Adult Literacy Survey au moyen d'échantillons de liaison ; et d) la création de tableaux de conversion des scores obtenus aux TALS.

Caractéristiques relatives aux échelles et aux tests

Probabilités de réponse. L'application du modèle IRT a notamment pour avantage de fournir de l'information sur les différents éléments des questionnaires qui permet de décrire les niveaux de rendement correspondant à chaque échelon de chaque échelle. Il devient possible de ce fait de déterminer les tâches que le sujet dont le score se situe à un niveau donné pourrait fort probablement accomplir avec succès, alors que le sujet dont le score se situerait à l'échelon juste en-dessous aurait des probabilités de succès bien moins élevées.

Relations inter-échelles. Les corrélations estimatives entre les trois tests mesurant le degré d'alphabétisation étaient respectivement de 0,59 (lecture de textes et recherche documentaire), de 0,60 (lecture de textes et calcul) et de 0,62 (recherche documentaire et calcul).

Rendement différentiel des groupes. Les stimulis retenus pour les tests ont été délibérément choisis pour refléter une vaste gamme d'intérêts et de contextes qu'on retrouve chez les adultes de façon qu'aucun groupe n'ait d'avantage par rapport à l'autre en raison de connaissances professionnelles très spécialisées. Par ailleurs, les questions ont été élaborées de manière à faire appel aux compétences liées aux divers types et niveaux de traitement correspondant aux trois domaines d'alphabétisation que l'on retrouve dans la vie de tous les jours.

BASIC ENGLISH SKILLS TEST
(test de compétences fondamentales en anglais)

Centre for Applied Linguistics
1118 nord-ouest, 22e Rue
Washington (DC) 20037
Tél.: 202-429-9292

Contenu

Le test de compétences fondamentales en anglais (BEST) a été conçu pour mesurer la capacité, de compréhension auditive, d'expression orale, de lecture et d'écriture correspondant au niveau fondamental et au niveau intermédiaire. Il se compose de deux volets, un volet entrevue orale et un volet degré d'alphabétisation, qui permettent de déterminer dans quelle mesure le sujet possède les compétences langagières fonctionnelles de base.

Le volet entrevue orale comprend 49 questions, réparties entre les domaines suivants : renseignements personnels, caractéristiques du foyer, heure, directions, achats et opérations financières simples, affiches, cabinet du médecin, accidents et formulaires d'emplois et de données. Le sujet doit aussi lire un passage et rédiger un texte. Ces deux tâches permettent de sélectionner les sujets à qui il y aurait peut-être lieu d'administrer la partie du test mesurant le degré d'alphabétisation.

Toutes les questions simulent diverses situations réelles de compréhension auditive et d'expression orale. Cette partie du test consiste en un entretien personnel d'une quinzaine de minutes.

Le volet degré d'alphabétisation comprend une cinquantaine d'items classés selon les rubriques suivantes : renseignements personnels, calendrier, étiquettes apposées sur les aliments et les vêtements, chèque de loyer, enveloppes, annuaire téléphonique, horaire de train, affiches, lecture d'un passage et rédaction d'un texte. Cette partie du test peut être administrée à un ou à plusieurs sujets et dure environ une heure.

Scores

Chacun des items du volet entrevue orale du BEST est conçu pour mesurer les compétences dans un des quatre domaines suivants : compréhension auditive, communication, maîtrise de la langue et lecture et écriture. Dans le guide d'accompagnement, chacun de ces quatre domaines est représenté par un symbole. Le guide contient une échelle de scores pour chaque item compris dans chaque domaine de compétences de même qu'un résumé des règles de notation pour

l'entrevue orale. On trouve dans le guide une courte description et des exemples qui correspondent à chacun des chiffres de l'échelle, qui va de zéro à trois. Une fois qu'il a fait le total des scores à l'aide d'une feuille de notation, l'évaluateur détermine le score total pour l'entrevue orale. Les sujets qui obtiennent un score de cinq ou plus peuvent être appelés à faire le test mesurant le degré d'alphabétisation.

Une feuille de notation du degré d'alphabétisation est utilisée pour ce volet du test. Un score de un ou de zéro est donné pour chacune des onze sections du test. Le guide d'accompagnement comprend des exemples et des instructions sur la façon de noter. Le score total pour l'ensemble des questions est inscrit sur la feuille de notation. Suivant l'approche axée sur les compétences, sept niveaux de rendement sont établis pour chacun des volets du test, chaque niveau correspondant à une échelle de scores différente.

L'information obtenue grâce aux deux volets du test peut être utilisée à des fins diverses. Les tests de classement ou d'admission évaluent la nature et l'étendue des compétences langagières des étudiants inscrits à un cours d'anglais. En classant les étudiants selon le score obtenu au test BEST, on peut obtenir des classes plus ou moins homogènes. Le test peut également servir à mesurer les

progrès des étudiants dans l'acquisition des compétences langagières fonctionnelles. Il peut aussi déterminer les progrès réalisés dans un cours en particulier à condition que le cours soit conçu pour enseigner les compétences langagières nécessaires pour survivre en société ou pour occuper un emploi. Le test peut également être utilisé à des fins diagnostiques. Le test BEST est conçu pour refléter le plus fidèlement possible les situations dans lesquelles les étudiants seront appelés à utiliser leurs compétences langagières pour pouvoir vivre de façon autonome. En examinant la performance de l'étudiant relativement à chacun des items du test, on obtient des informations importantes quant à sa capacité à accomplir des tâches langagières données. Enfin, le test peut être utilisé à des fins de sélection, notamment pour déterminer si l'étudiant est en mesure de suivre un programme d'études ou de formation donné.

Processus d'élaboration

Les spécifications à partir desquelles le test a été élaboré mettaient l'accent sur les domaines où il fallait avoir un niveau de compétence minimal en anglais pour vivre en société. Le nombre total d'étudiants qui ont participé aux essais sur

le terrain aux États-Unis des formulaires B, C et D étaient respectivement de 987 pour le volet entrevue orale et de 632 pour le volet degré d'alphabétisation. Le choix des items à inclure dans le formulaire qui a été retenu pour le test s'est fait essentiellement à partir des résultats statistiques de l'analyse des différents items. Chaque item a été examiné au regard du niveau de difficulté et de la corrélation bisériale (coefficients r). L'analyse a révélé que la plupart des items des formulaires mis à l'essai pouvaient être conservés. La fiabilité estimative au regard de la cohérence interne des trois formulaires allait de 0,77 à 0,83 pour ce qui est des sous-scores et de 0,96 à 0,97 pour les scores totaux. Ces scores sont très élevés compte tenu du fait que le test est composé en grande partie de questions ouvertes. L'analyse de la fiabilité inter-évaluateurs, qui compare les scores que les différents évaluateurs donnent à un même sujet, a révélé un degré d'uniformité élevé entre les notes des évaluateurs.

La validité apparente du test BEST était aussi élevée puisque son contenu reflète assez fidèlement l'accomplissement en situation réelle des tâches langagières qu'il vise à mesurer. L'analyse de la corrélation entre les scores obtenus et ceux qui avaient été attribués au groupe au départ pour la performance langagière globale des sujets permet de conclure à la validité conceptuelle du test.

Pour ce qui est des corrélations entre les sous-échelles, chaque sous-échelle du volet entrevue orale du test témoigne d'une importante corrélation positive avec le score total pour ce volet. De même, il y a une étroite corrélation entre les scores de lecture et d'écriture pour les trois formulaires.

CANADIAN ADULT ACHIEVEMENT TEST
(test canadien de rendement pour adultes)

The Psychological Corporation
Harcourt Brace and Company
555, av. Horner
Toronto (ON) M8Z 4X6
Tél.: 1-800-387-7278

Contenu

Le test canadien de rendement pour adultes (CAAT) est destiné à mesurer le niveau scolaire atteint par un sujet adulte. Il s'agit d'une batterie de sous-tests en vocabulaire, en compréhension de lecture, en orthographe, en calcul arithmétique, en résolution de problèmes, en raisonnement mécanique, en langue, en sciences et en techniques d'étude. Le CAAT comporte quatre niveaux : le niveau A est destiné aux adultes qui ont de un à trois ans d'instruction, le niveau B, aux adultes qui ont de quatre à six ans d'instruction, le niveau C, aux adultes qui ont de sept à dix ans d'instruction, mais qui n'ont pas nécessairement terminé leurs études secondaires, et le niveau D, aux adultes qui ont au moins onze ou douze ans d'instruction et qui ont probablement terminé leurs études secondaires. Les différents sous-tests ne se retrouvent pas tous aux quatre niveaux.

Le CAAT évalue et décrit les compétences génériques de l'apprenant dans l'accomplissement de différentes tâches : décoder des mots, lire des textes de nature scolaire, repérer les fautes d'orthographe, faire des déductions à partir de différents types de textes, faire des additions et des soustractions et savoir résoudre des problèmes liés à la consommation. Les paragraphes ci-dessous donnent au lecteur une idée du contenu de certains des sous-tests.

Le sous-test de vocabulaire est conçu pour évaluer la connaissance et la compréhension de mots que les adultes rencontrent souvent au travail ou dans la vie quotidienne. Le sous-test comprend 34 questions à choix multiples concernant des mots tirés du vocabulaire appliqué ou général, du vocabulaire des sciences physiques et naturelles et du vocabulaire des sciences sociales. Le sous-test de résolution de problèmes comporte 21 problèmes et mesure la capacité du sujet à trouver une solution, à consigner et à récupérer de l'information et à mesurer à l'aide de notions de géométrie. L'apprenant est appelé à résoudre des problèmes comportant des entiers, des décimales, des pourcentages et des fractions du genre de ceux qui se présentent fréquemment dans la vie d'un adulte. Le sous-test de langue, qui est tiré du Stanford Test of Academic Skills (test Stanford de compétences scolaires), est conçu pour mesurer la connaissance fonctionnelle et

l'utilisation efficace de la langue anglaise. Le sous-test comprend 54 questions, réparties en quatre sections : recherche documentaire, sensibilité langagière, conventions et disposition en paragraphes.

Scores

Le test CAAT permet d'obtenir deux types de scores : des scores liés au contenu et des scores normatifs. La première catégorie comprend notamment le score brut objectif. Les scores bruts moyens pour tous les objectifs de chaque niveau du CAAT ont été déterminés pour l'échantillon à qui chaque niveau du test a été administré dans le cadre du programme de recherche sur le CAAT. La performance individuelle ou collective relativement à chacun des objectifs du CAAT peut être évaluée par rapport à ces scores moyens afin de déterminer les points forts et les faiblesses du ou des sujets.

Les scores normatifs servent à comparer la performance du sujet à celle d'un groupe de référence comparable. Les équivalents scolaires et les scores gradués sont des scores normatifs du genre de ceux qui ont déjà été utilisés pour mesurer la performance au CAAT. L'équivalent scolaire est déterminé en

comparant le score du sujet à la performance typique d'un certain nombre de groupes d'adultes à qui le test a été administré et qui se trouvaient à un niveau scolaire donné à un mois donné de l'année scolaire. Les équivalents scolaires du CAAT vont du jardin d'enfants à 12,9 années d'instruction, après quoi on passe au niveau postsecondaire. Malgré les limites qu'ils présentent, les scores fondés sur l'équivalent scolaire peuvent donner une indication du niveau approximatif auquel il faudrait débuter l'enseignement.

Les scores gradués du CAAT représentent la performance aux sous-tests et les totaux pour les quatre niveaux. Ainsi, étant donné que le sous-test de compréhension de lecture se retrouve aux trois niveaux du CAAT, les scores gradués pour ce sous-test sont comparables, qu'il s'agisse du niveau A, B, C ou D. À cause du caractère continu de l'échelle et du fait qu'elle est divisée en intervalles égaux, les scores gradués conviennent tout particulièrement à l'analyse des progrès réalisés pendant la période d'enseignement, peu importe le niveau du CAAT qui a été administré avant et après l'enseignement.

Processus d'élaboration

Parce que le CAAT vise à mesurer le rendement scolaire d'adultes, on a jugé essentiel que l'étalonnage se fasse, non pas avec des enfants d'école, mais avec des adultes. Cependant, malgré tous les travaux consacrés ces derniers temps à la mesure du rendement scolaire des adultes, il n'existe toujours pas de critères acceptables pour définir la population d'adultes canadiens à qui le CAAT conviendrait. Étant donné l'absence de ces critères, la recherche a porté sur un certain nombre de groupes d'adultes dont les caractéristiques serviraient à définir les utilisateurs de cet instrument. L'échantillon a été établi à partir de listes d'adultes inscrits à des programmes scolaires et professionnels dans divers établissements communautaires et scolaires et à partir de listes d'envoi d'établissements pénitentiaires provinciaux et régionaux. La recherche a permis d'obtenir des informations quant au bien-fondé des questions et des objectifs, quant à la difficulté des questions et à la façon dont elles tenaient compte de la réalité adulte et quant à l'efficacité des options données pour toutes les questions à choix multiples.

Fiabilité. La fiabilité du test est fonction de la mesure dans laquelle il donne systématiquement les mêmes résultats. La fiabilité des scores du CAAT a été déterminée à l'aide de la formule Kuder-Richardson 20. Les estimations de fiabilité vont de 0,84 à 0,95 pour les niveaux A, B et C.

Validité. Pour évaluer la validité du contenu du texte, il faut déterminer s'il mesure les compétences fondamentales dont l'adulte a besoin pour fonctionner dans notre société moderne. Étant donné que les objectifs ont été rédigés de manière à refléter le contenu des programmes d'éducation des adultes à l'échelle du Canada, le CAAT est un instrument valable pour mesurer ces objectifs communs. Il appartient toutefois aux responsables des différents programmes de déterminer la validité du CAAT pour mesurer leurs objectifs. Par ailleurs, l'analyse corrélationnelle a révélé que les tests du CAAT témoignent d'une corrélation positive les uns avec les autres.

MEASUREMENT, READING AND ARITHMETIC
(mesure, lecture et calcul)

Ramsey Corporation
Boyce Station Offices
1050, ch. Boyce
Pittsburg (PA) 15241-3907
Tél.: 412-257-0732

Contenu

Ce test critériel porte sur trois compétences fondamentales, la mesure, la lecture et le calcul. La mesure d'entiers, de moitiés, de quarts, de huitièmes et de seixièmes occupe la première partie du test destinée à des stagiaires qui doivent savoir mesurer avec exactitude. Cette partie du test se compose de 20 items à choix multiples. La partie lecture du test comprend cinq paragraphes exigeant un niveau de lecture de plus en plus élevé. Chacun des cinq passages est suivi de huit questions à choix multiples. La compréhensibilité estimée des passages en question va de 8,0 dans le cas des instructions à 12,0 dans le cas du dernier passage. Le niveau moyen de compréhensibilité est de 10,6.

Les passages liés au travail sont destinés à mesurer la capacité à lire des informations liées au travail du genre de celles qu'on trouve dans divers domaines

industriels : fluides hydrauliques, outils d'établi, hydraulique, lubrification et solénoïdes. La partie calcul du test est conçue pour évaluer la capacité à réaliser des opérations arithmétiques simples, additions, soustractions, multiplications et divisions, avec des entiers, des décimales et des fractions. Elle comprend aussi des tableaux et des formules simples. Le test se compose de 40 questions à choix multiples. Chaque partie du test est d'une durée limitée. Pour la mesure, la durée est de 15 minutes ; pour la lecture, elle est de 40 minutes ; et pour le calcul, elle est de 30 minutes.

Scores

Il s'agit d'un test critériel où un seuil ou un critère de réussite absolu est établi et le score du sujet est déterminé par rapport à ce seuil. Ainsi, un score de 90 p. 100 ou plus correspondrait à un A, un score de 80 à 89 p. 100 correspondrait à un B et ainsi de suite. Des équivalents centiles ont été établis pour les trois parties du texte à partir des scores obtenus par un échantillon de 1 441 métallurgistes.

Processus d'élaboration

Des études de fiabilité ont été réalisées sur d'autres tests de la Ramsey Corporation, comme le Technician Electrical Test, le Technician Mechanical Test, le Building Maintenance Test, l'Electronics Test et l'Electrical Maintenance Trainee Test, mais aucune n'a été faite sur ce nouvel instrument. Sept étapes ont toutefois été prévues pour assurer la validation fondamentale du test. Premièrement, le client fournit une description du travail qui se fait dans son entreprise, laquelle description est reprise suivant les rubriques retenues par la Ramsey Corporation, énoncé des fonctions, travail effectué et conséquences des erreurs. Dix spécialistes du domaine complète la description pour faire en sorte qu'elle soit à jour et qu'elle reflète bien les fonctions. Puis, les dix spécialistes évaluent le travail en question au regard de l'importance des différentes fonctions et du temps en pourcentage qui y est consacré, et leurs évaluations sont ensuite confrontées. La troisième étape consiste à évaluer les niveaux de compréhensibilité, les compétences mathématiques et les autres exigences du travail et du programme de formation. L'entreprise chargée de dresser le test évalue ensuite l'information obtenue de l'analyse du travail et détermine les connaissances, les compétences et les capacités requises de même que les

méthodes de mesure. La cinquième étape comprend la révision, la mise en pages et l'impression du test. Le test est ensuite soumis à un groupe d'au moins quatre spécialistes du domaine pour déterminer s'il reflète bien les connaissances, les compétences et les capacités visées. Puis, on y applique la procédure Angoff pour établir les notes de passage et assurer la révision définitive. La révision est faite par un groupe de cinq à dix spécialistes du domaine qui répondent eux-mêmes aux différents items du questionnaire et qui comparent ensuite leurs réponses avec la bonne réponse pour déterminer quel est le pourcentage de candidats qualifiés qui auraient la bonne réponse. Une fois les indices de difficulté et de la corrélation bisériale de point calculés, la dernière étape consiste à rédiger un rapport sur la validation du contenu suivant la présentation prévue dans l'Uniform Guideline on Employee Selection Procedures (lignes directrices uniformes sur les procédures de sélection des employés). L'entreprise chargée de dresser le test examine le projet de rapport et envoie au client le rapport définitif décrivant dans le détail le processus de validation suivi.

WIDE RANGE ACHIEVEMENT TEST 3
(test de rendement général 3)

Jastak Associates
C.P. 3410
Wilmington (DE)
19804-0250
Tél.: 1-800-221-9728

Contenu

Le test de rendement général 3 (WRAT3) est destiné à mesurer les codes nécessaires pour acquérir les compétences fondamentales en lecture, en orthographe et en arithmétique. Cette version récente du test peut être administrée à des adultes de 17 à 75 ans. Chacun des deux formulaires pouvant être utilisés comprend trois sous-tests. Le sous-test de lecture se compose de deux sections, une où le sujet est appelé à lire des lettres et l'autre où il doit lire des mots. Dans la section portant sur les lettres, le sujet est appelé à nommer 15 lettres de l'alphabet, tandis que dans la section portant sur les mots, il est appelé à prononcer 42 mots. Les deux sections du sous-test de lecture doivent être administrées individuellement. Dans le cas des adultes, la section portant sur les mots est administrée en premier. Il s'agit, non pas d'un test de prononciation ou de diction, mais bien d'un test de reconnaissance des mots. Quand les circonstances exigent

un test de lecture plus complet, les deux sections du test peuvent être administrées. Un point est accordé pour chaque lettre et chaque mot lu correctement. Les scores bruts sont inscrits sur le questionnaire dans l'espace prévu à cet effet.

Le sous-test d'orthographe comprend des lettres de l'alphabet et des mots qui sont dictés aux sujets. Les deux sections du test se composent de 55 items en tout : 15 lettres de l'alphabet et 40 mots. Quand il s'agit d'adultes, la section portant sur l'orthographe des mots est administrée en premier. Les mots que la personne chargée d'administrer le test doit dicter sont écrits sur des cartes de plastique avec des indications sur la façon de les prononcer. Un point est accordé pour chaque lettre et chaque mot bien écrit. Les scores bruts d'orthographe sont ensuite inscrits sur le questionnaire dans l'espace prévu à cet effet.

Pour le sous-test d'arithmétique, le sujet doit compter, lire des symboles numériques, résoudre des problèmes qui lui sont exposés oralement et faire des calculs par écrit. Le sous-test se divise lui aussi en deux sections, une section orale comportant 15 items et une section écrite comportant 40 items. Les adultes commencent par la section écrite. Chaque élément compte pour un point et le score brut du sous-test d'arithmétique est inscrit sur le questionnaire dans l'espace

prévu à cet effet. Chacun des sous-tests du WRAT3 est d'une durée de 15 à 30 minutes.

Scores

Plusieurs types de score sont utilisés pour consigner les résultats obtenus au WRAT3. Quatre types de scores sont décrits ci-dessous. Les scores bruts correspondent au nombre total de bonnes réponses. Ils ne peuvent pas donner une mesure précise d'une performance par rapport à une autre. Les scores absolus, qu'il s'agisse d'un seul sujet ou d'un groupe de sujets, mesurent le degré de réussite pour chaque variable quelle qu'elle soit et sans égard au niveau scolaire ou à l'âge. Ces scores conviennent surtout pour les tests administrés avant et après un programme d'enseignement. Les scores normatifs sont ceux qui sont utilisés pour comparer les sujets à eux-mêmes et à d'autres. Il est ainsi possible de déterminer non seulement l'écart entre deux scores, mais aussi l'importance de cet écart. Voici des exemples d'évaluations faites à partir de scores normatifs : moyen supérieur (110-119), moyen (90-109), moyen inférieur (80-89). Les scores correspondant au niveau scolaire donnent une évaluation un peu plus précise que

les scores bruts. Dans le cas du WRAT3, ces scores ne permettent de se faire qu'une idée approximative du niveau scolaire atteint en raison de la variation considérable qu'on constate à l'intérieur d'un niveau scolaire, tout comme entre les différents niveaux scolaires. Ces scores devraient être utilisés avec l'information obtenue à partir des scores normatifs ou des scores absolus pour décrire les résultats du test.

Processus d'élaboration

Le WRAT3 est un test normatif fondé sur l'âge qui mesure les capacités de codage scolaire suivantes : reconnaissance de mots, orthographe et calcul arithmétique. Une nouvelle normalisation de cette version du test a été entreprise au début des années 90 à la suite de l'analyse Rasch du degré de difficulté des éléments du test. Cette analyse avait révélé de nombreux cas de double emploi dans les items des deux formulaires du WRAT-R. Le WRAT3 a été administré à un échantillon étalonné aux États-Unis. L'échantillon adulte a été sélectionné à partir d'un modèle d'échantillonnage stratifié avec ajustement en fonction de l'âge, de la région du domicile, du sexe, de l'origine ethnique et du niveau socio-

économique. Au total, 4 834 personnes ont été échantillonnées pour tous les groupes d'âge (5-75).

Fiabilité. Quand on mesure une compétence, on devrait normalement pouvoir s'attendre à ce qu'une autre mesure de la même compétence donne des résultats semblables. La fiabilité du test est fonction de la mesure dans laquelle il permet d'évaluer une compétence en particulier de façon systématique. Il existe de nombreuses mesures de la fiabilité d'un test. Voici les résultats des trois mesures qui ont été appliquées au WRAT. Les coefficients alphas médians des différents formulaires du WRAT vont de 0,85 à 0,95, ce qui est considéré comme souhaitable. Les coefficients de corrélation des scores bruts des trois sous-tests sont aussi très élevés : 0,98, 0,98 et 0,98 respectivement. Selon la méthode test-retest, les coefficients de stabilité corrigés vont de 0,91 à 0,98.

Validité du contenu. Les statistiques Rasch permettant d'analyser chaque item ont permis de déterminer dans quelle mesure les items choisis étaient représentatifs des trois domaines de compétence : décodage en lecture, encodage en lecture et calcul arithmétique. Pour chacun des sous-tests, on a constaté que le score obtenu à cet égard était le plus élevé possible, soit 1,00, ce qui tend

fortement à confirmer la validité du contenu à l'égard de chacune des trois variables.

Validité conceptuelle. Les données concernant la nature évolutive du WRAT3 révèlent un accroissement constant des scores moyens au fur et à mesure qu'on passe d'un groupe d'âge étalonné à un autre, jusqu'à ce qu'on atteigne les groupes ayant entre 45 et 54 ans. À partir de là, les scores commencent à tomber. Les données relatives à l'ensemble de l'échantillon étalonné permettent de conclure à l'existence d'une corrélation significative entre le score brut au test WRAT3 et l'âge en années et en mois du sujet. L'analyse des corrélations indique par ailleurs que les formulaires du WRAT3 ont une corrélation positive les uns avec les autres. Ainsi, la corrélation entre le test de lecture et le test d'orthographe va de 0,81 à 0,91 ; la corrélation entre le test de lecture et le test d'arithmétique va de 0,54 à 0,78 et la corrélation entre le test d'orthographe et le test d'arithmétique va de 0,58 à 0,82. Ces corrélations indiquent un rapport allant de modérément à fortement élevé entre les variables.

METRIC SKILLS ASSESSMENT
(évaluation des compétences métriques)

Lakeshore Technical College
Gary Heeler
1290, North Avenue
Cleveland (WI) 53015
Tél.: 414-458-4183

Contenu

L'évaluation des compétences métriques est un test axé sur les compétences qui est conçu pour mesurer les résultats d'une série préétablie d'expériences d'apprentissage métrique. Le test, qui comprend 35 items, est divisé en deux sections. La première section porte sur l'origine, les unités fondamentales de mesure, les définitions et les conversions numériques du système métrique. La deuxième section comprend une série de questions qui amènent le sujet à consigner les mesures obtenues à l'aide de différents instruments de mesure utilisés au travail, comme le micromètre et le pied à coulisse. Le test peut être administré à un ou plusieurs apprenants adultes, et il est d'une durée de 20 minutes environ. Le test peut par ailleurs être utilisé pour l'évaluation préalable ou postérieure à la formation.

Scores

Le test Metric Skills Assessment est un outil critériel, de sorte que les progrès accomplis par l'apprenant sont interprétés en fonction du nombre d'unités d'instruction qu'il termine avec succès au niveau critériel prescrit. Le test peut être administré préalablement à un programme de formation et, selon le critère utilisé par l'instructeur, il peut déterminer la composante du programme d'études qui convient à l'apprenant. Quand le test est utilisé à la fin d'un programme de formation, le critère préétabli se révèle généralement correct à 90 % ou 80 %.

Processus d'élaboration

La première étape dans l'élaboration du test a été de déterminer avec l'aide d'un groupe de spécialistes des mathématiques les principales compétences qui devaient être acquises. Puis, on a repéré divers matériels pédagogiques offerts sur le marché qui présentaient une multitude d'exercices d'apprentissage permettant d'acquérir chacune des compétences visées. Autrement dit, on a d'abord défini les compétences à acquérir, puis on a conçu le programme d'études de façon à aider

l'apprenant à acquérir ces compétences. Le programme d'études comprend six domaines clés. Le premier module consiste en une initiation aux unités de mesures métriques qui sont utilisées dans la vie quotidienne de même qu'une révision des opérations décimales. Ce module comprend les mesures traditionnelles et métriques de même que l'utilisation d'échelles et de galons métriques. Le deuxième module porte sur la conversion entre les unités de mesure métriques et britanniques, tandis que le troisième comprend le filetage métrique et l'identification des appareils de mesure métrique. Dans le quatrième module, on apprend à identifier et à utiliser les instruments de précision, tandis que dans le cinquième, on apprend à utiliser le pied à coulisse. Le programme se termine par une révision du système de mesure métrique et par une réévaluation des compétences.